# NOTES 23: INFERENCE FOR >2 MEANS

Stat 120 | Fall 2025
Prof Amanda Luby

A biology student tested four different hand-washing methods (water only, washing with regular soap, washing with antibacterial soap, and spraying hands with antibacterial spray) and counted the number of bacteria present after two days of growth. They did this eight times for each of the four treatments.
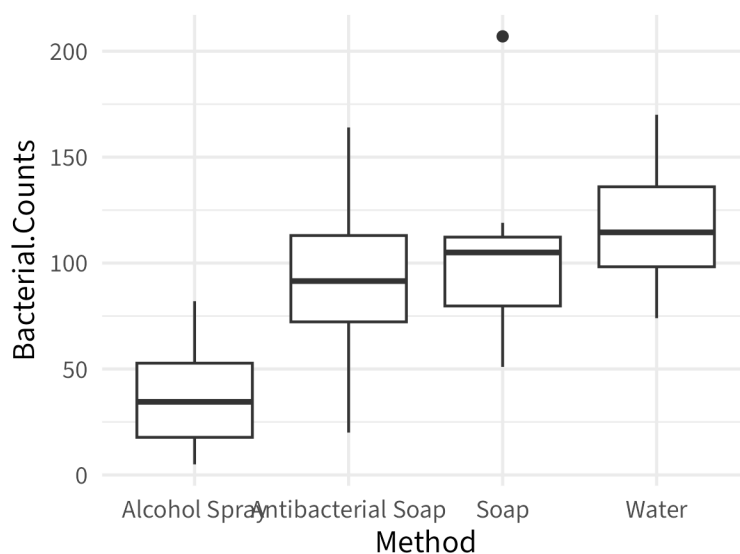
Up until now, we've done something like the following:

H_0 :

H_A :

What are some potential issues?

1 ANOVA

ANOVA stands for **ANalysis Of VAriance**, and it's a method we use to compare means of groups when we have more than two groups.

Here, the null hypothesis we are interested in is:

H_0 :

H_A :

But, it turns out it's much harder to test multiple means than comparing only two.

Idea: (Visual)

Our test statistic is called an "F" statistic and compares the variance within the groups to the variance across the groups.

$$F = \frac{\text{Sample variance of means}}{\text{Pooled estimate of variance within groups}} = \frac{MS_T}{MS_E}$$

In R, we calculate the F-statistic using `aov(Response ~ Explanatory)` and then `summary()`. This is the same syntax that we used for linear regression (`lm()`), we're just using a categorical variable instead of a quantitative variable as our predictor.

```
          Df Sum Sq Mean Sq F value Pr(>F)
Method     3  29882    9961    7.06 0.0011 **
Residuals 28  39484    1410
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This is called an ANOVA table and was originally designed to organize the calculations needed to compute an F-statistic.

> *This table has a long tradition stretching back to when ANOVA calculations were done by hand. Major research labs had rooms full of mechanical calculators operated by women. (Yes, always women) Three women would perform each calculation, and if any two of them agreed on the answer, it was taken as the correct value.*

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Method	k-1	29882	SumSq/df	MS_t/MS_e	
Residuals	N-k	39484	SumSq/df		

1.1 Assumptions

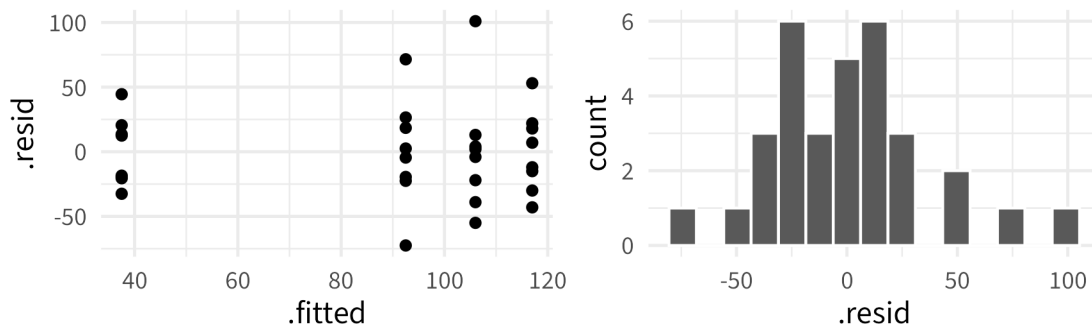We need to verify some assumptions whenever we fit an ANOVA model.

1. Independence

Have to check based on the data collection process

2. Equal Variance

- Do the boxplot "whiskers" look roughly the same length across all groups?
- Check the residual vs fitted plot:

3. Normal residuals:

3

- Check histogram of residuals

2 Post-hoc Testing

So far we've introduced ANOVA as a method for comparing many means to each other concurrently. Finding a statistically significant result at the end of an ANOVA however only tells us that at least one pair of means are different, but not *which pair(s)* of means are different.

We won't do a thorough discussion of post-hoc testing in this class, but the R code to do all pairwise tests is:

```
pairwise.t.test(x = Soap$Bacterial.Counts, g = Soap$Method, p.adjust.method = "none")
```

```
                  Alcohol Spray Antibacterial Soap    Soap
Antibacterial Soap    0.0066862                 NA      NA
Soap                  0.0010697             0.4781      NA
Water                 0.0002236             0.2026  0.5627
```

This approach suffers from a **multiple testing problem**. We can adjust the p-values to get valid results:

```
pairwise.t.test(x = Soap$Bacterial.Counts, g = Soap$Method, p.adjust.method = "holm")
```

```
                  Alcohol Spray Antibacterial Soap    Soap
Antibacterial Soap     0.026745                 NA      NA
Soap                   0.005349             0.9562      NA
Water                  0.001342             0.6077  0.9562
```

4